



# OAI-PMH : Alma, Primo, Rosetta and more

22 March 2023

---

# Introduction

[Deborah Fitchett](#)

Associate University Librarian  
Lincoln University





## What is OAI-PMH?

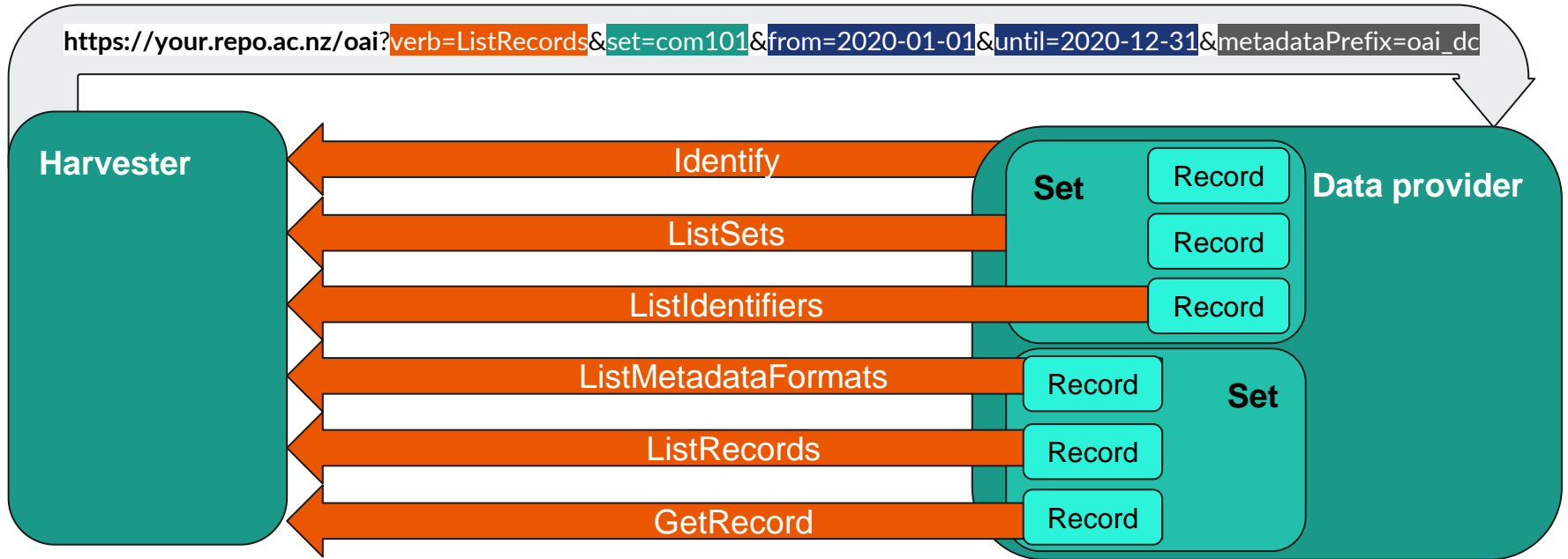
**O**pen  
**A**rchives  
**I**nitiative  
**P**rotocol for  
**M**etadata  
**H**arvesting

So it's an open access standard which describes a way to let metadata be harvested from a data provider.

More: <https://www.openarchives.org/pmh/>

- Includes a link to full specifications for the technically minded!

# Harvester requests - data provider responds





# Examples

## Harvesters

- Primo
- DigitalNZ
- Trove
- OAlster
- ANBD
- ....

## Data providers

- Research repositories (eg DSpace / ePrints)
- Heritage sites (eg ContentDM, Recollect, etc)
- Special collections (eg Figshare, Open Journal Systems, Equella)
- Rosetta, Alma-D, Alma
- Wordpress (with custom code...)
- ....



# Today's presentations

## Trove

[Witteveen](#) (La Trobe) and [Peta Hopkins](#) (Bond)

- [David](#)

## Harvesting

- Figshare -> Primo
- Rosetta -> Alma University)

- [David Witteveen](#) (La Trobe)  
- [Mike Kmiec](#) (Western Sydney)

## Publishing

- RapidILL, Recollect -> Primo
- Alma -> NLNZ search & Rosetta

- [Bronwyn King](#) (UNSW)  
- [Rich Robertson](#) (NLNZ)



Trove

---

# Setting up OAI-PHM harvesting for Trove

[David Witteveen](#)

Coordinator, Library Discovery Platforms  
La Trobe University Library





---

## The goal

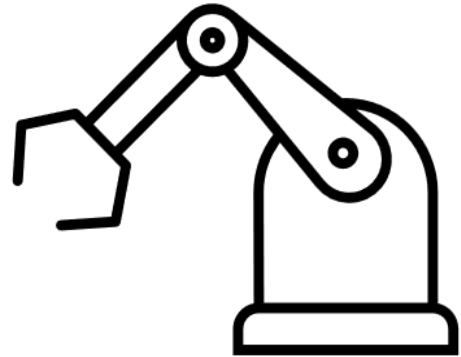
Allow Libraries Australia to automatically harvest our physical holdings from Alma and add them to Trove.

### Why?

Make our collections discoverable for resource sharing requests.

### Before:

Our Collections team manually created a file of our holdings once a week, and then FTPed it to Trove.





## The process - in theory





## The process - reality

It took longer than expected.


We got busy. Trove got busy. We missed steps. We forgot who was waiting on who.

I also got Long COVID.

Read the [instructions from Ex Libris](#)

AUG	<p>Trove sent us their questionnaire. Collections team filled it out, and sent it back.</p> <p>About a three week wait for a follow-up from Trove</p>
SEP	<p>Trove told us they were ready to set up harvesting. We needed to provide them with the details of the set.</p>
OCT	<p>We finished configuring the set in Alma. Sent to Trove.</p>
NOV	<p>Trove tested our set. Couldn't harvest!</p> <p>I needed to make some changes in Alma:</p> <ul style="list-style-type: none"><li>- Added Trove's IP range to the OAI integration</li><li>- Created a Publishing Profile</li></ul>
DEC	<p>Initial harvest. Took 8 days (!!!).</p>
JAN	<p>Clean up:</p> <ul style="list-style-type: none"><li>● Fixed an issue with deep linking</li><li>● Removed some duplicate holdings under our old NUCs</li><li>● Withdrew our old NUCs</li><li>● Changed how La Trobe University is listed</li></ul> <p><b>DOCUMENT THE NEW PROCESS!!!</b></p>

# Instructions from Ex Libris



The screenshot shows the Ex Libris Knowledge Center interface. At the top, there is a navigation bar with "Ex Libris Knowledge Center" and "Part of Clarivate" on the left, and "More Sites", "System Status", and "Submit a Case" on the right. Below this is a breadcrumb trail: "Home > Alma > Community Knowledge > Publish MARC records for digital inventory for the Australian". The main heading is "Publish MARC records for digital inventory for the Australian National Bibliographic Database and Trove using the OAI-PMH daily update". Below the heading, it says "Created By: Peta Hopkins" and "Created on: 10/02/2020". There is a note: "Co-authored with Amelia Rowe from RMIT University". The section "Record considerations" includes the text: "Consider the National Electronic Deposit standards (Australia) in preparing your records." The section "Create a set in Alma" lists four numbered steps: 1. To limit the set for digital resources select "Digital Titles" in the scope. However, if you plan to use a single set for publishing different inventory types, then you would need to choose "All Titles". 2. Optionally you may choose to use the Tag Sync National Catalog option - this will depend on your cataloguing processes, and if you want to emulate the same process used for the Libraries Australia Publishing Profile 3. Choose Record Format = MARC21 4. Consider using a naming convention for your sets as they can be hard to find if you have a lot at your institution.

## Publish MARC records for digital inventory for the Australian National Bibliographic Database and Trove using the OAI-PMH daily update

1. Create a set in Alma
2. OAI integration - allow Trove's IP addresses
3. Create a publishing profile
4. Talk to Trove

(Note the credits!)

# The End Result

We now have a neat, clean listing in Trove.

Our holdings are updated every night.

## STILL TO DO:

- We need to automate adding items to the set
- A bug in Trove mean we're not getting harvest report emails.

## The politics and reality of family care in corporate America / John P. Fernandez

2 Editions under this title



Read ▾ Borrow (6) ▾ Buy ▾

ALL 6 ACT 1 NSW 1 QLD 1 VIC 2 WA 1

Charles Sturt University. Bathurst Campus Library. <a href="#">🔗</a> Book; Illustrated	<b>La Trobe University, La Trobe University Library. <a href="#">🔗</a> Book; Illustrated</b>	The University of Melbourne, The University of Melbourne Library. <a href="#">🔗</a> Book; Illustrated
The University of Queensland, University of Queensland Library. <a href="#">🔗</a> Book; Illustrated	University of Western Australia, University of Western Australia Library. <a href="#">🔗</a> Book; Illustrated	



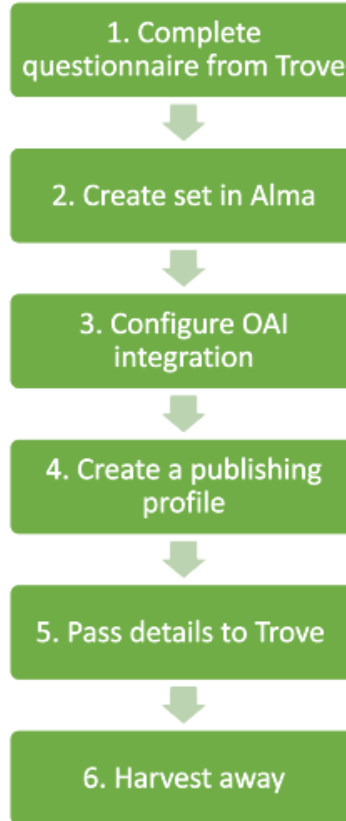
## Conclusion

### Advice:

- Allow time
- Read the [instructions from Ex Libris](#)
- You can do the work in Alma beforehand
- DOCUMENT YOUR PROCESS!

### Bonus:

- Everything I learnt in this project will be useful for setting up RapidILL





# Digital resources, and then...

## Since 2015

Publishing print holdings via RIS - Record Import Service - using the Alma Libraries Australia Publishing profile, a weekly FTP.  
Electronic holdings (just for Bond theses) added in later

## 2018

Started using Alma Digital is for our small, but growing collection of digitisations, mostly from the Bond University Archives - unique resources.

## 2019

Requested Ex Libris add digital resources to the Libraries Australia publishing profile (for RIS)  
Met & corresponded with Ex Libris, RMIT, UWA, Trove Collaborative Services (TCS) and others to work through the use of a general publishing profile instead.

- General publishing profile was preferred by Ex Libris
- OAI-PMH was preferred by TCS - just as well because the Alma general publishing profile did not support the file naming needed for RIS ftp method

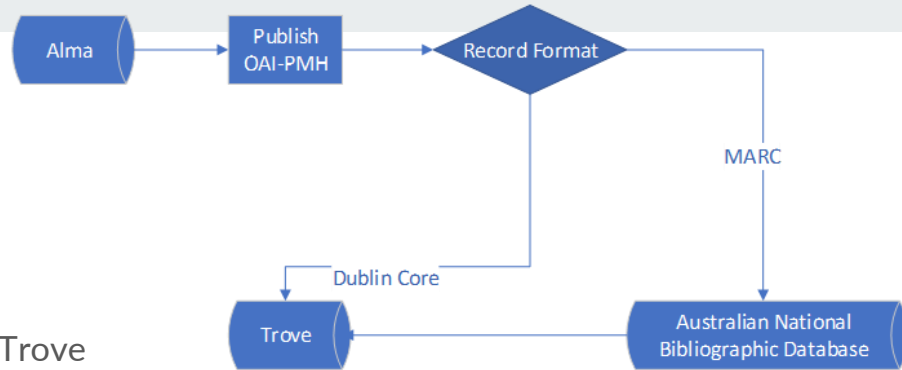
## 2020

Commenced publishing/harvesting  
RMIT, Bond & TCS presented at IGELU and then updated at ANZREG 2020 virtual conference [Slides](#) - [Recording](#)



**Peta Hopkins**  
Manager, Digital Library Services  
Bond University





## Record format - MARC 21

- Records harvested into ANBD and surfaced in Trove
- Data enrichment includes delivery and thumbnail URLs
- Set defined as Access rule = Open Access & Record format = MARC

```

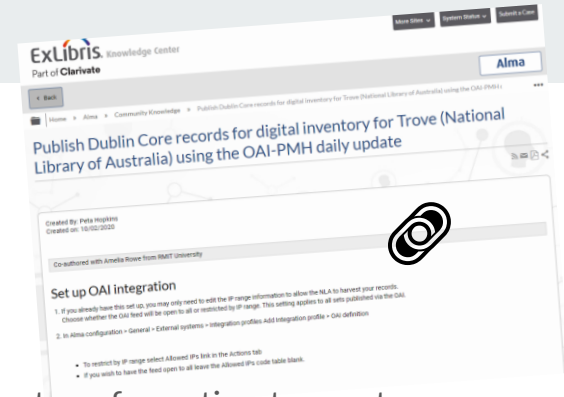
<subfield code="1">BOND COLLECTIONS</subfield>
</datafield>
▼<datafield tag="AVD" ind1="0" ind2=" " >
  <subfield code="b">128082990002381</subfield>
  <subfield code="f">Educational use permitted</subfield>
  <subfield code="e">Proposed site for Bond University</subfield>
  <subfield code="s">1 file/s (jpg)</subfield>
  <subfield code="m">2018-11-12 01:47:05</subfield>
  <subfield code="g">MAIN</subfield>
  <subfield code="i">VIEW</subfield>
  <subfield code="d">https://bond.alma.exlibrisgroup.com/view/delivery/thumbnail/61BOND_INST/128082990002381</subfield>
  <subfield code="n">2022-01-11 05:58:59</subfield>
  <subfield code="k">System</subfield>
  <subfield code="l">System</subfield>
  <subfield code="c">https://bond.alma.exlibrisgroup.com/discovery/delivery/61BOND_INST:BOND/128082990002381</subfield>
</datafield>

```



# Record format - Dublin Core

- Records harvested only into Trove
- Output only includes the delivery URL as an identifier - TCS makes a transformation to create a thumbnail URL based on the MMS ID in one of the identifiers
- Set defined as Access rule = Open Access & Record format = DC



```
<record>
  <header>
    <identifier>oai:alma.61BOND_INST:9930460592602381</identifier>
    <timestamp>2022-11-12T14:00:33Z</timestamp>
    <setSpec>LA-digital-dc</setSpec>
  </header>
  <metadata>
    <?oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:dcterms="http://purl.org/dc/terms/"
      xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
      <dc:title>Dave & Trish Moffatt</dc:title>
      <dc:creator>Dr Jane E. Hunt</dc:creator>
      <dc:subject>Triathlon</dc:subject>
      <dc:description>Part of a series of audio interviews held with people associated with the sport of triathlon in Australia. Dr Hunt used these interviews in her research for the book "Multisport dreaming: The foundations of
      Australia" published in 2014. Length of interview: 38:51</dc:description>
      <dc:date>14 April 2012</dc:date>
      <dc:type>Sound</dc:type>
      <dc:format>MP3</dc:format>
      <dc:format>37.31 MB</dc:format>
      <dc:identifier>alma:61BOND_INST/bibs/collections/8175669670002381</dc:identifier>
      <dc:identifier>https://bond.alma.exlibrisgroup.com/discovery/delivery/61BOND_INST:BOND/1282049690002381</dc:identifier>
      <dc:language>English</dc:language>
      <dc:relation>Audio Interviews</dc:relation>
      <dc:relation>alma:61BOND_INST/bibs/collections/8175669670002381</dc:relation>
      <dc:rights>This work is licensed under a Creative Commons <a href="https://creativecommons.org/licenses/by-nc-nd/4.0/">Attribution-Non Commercial-No Derivatives (CC BY-NC-ND) licence</a>.</dc:rights>
      <dc:rights>© Copyright, Jane Hunt, 2012</dc:rights>
      <dc:rights>Unrestricted online access</dc:rights>
    </oai_dc:dc>
  </metadata>
</record>
```

# Consolidating MARC records

2022 - 2023 - Moving all inventory types into a single publishing profile for MARC 21 records

Physical

Bulk of titles



Electronic

Bond University  
Theses



Digital

Bond University  
Archives

## Single publishing profile

Challenge: how to define the set(s) - max 2 sets  
Normalisation rules work across all inventory types - be careful

Search in: All titles ▾ Zone:  Institution  Community

Tag Sync National Catalog ▾	Equals ▾	Publish Bibliographic records ▾		
<b>AND</b> OR	Tag Suppressed (Title) ▾	Equals ▾	No ▾	
<b>AND</b> OR	Record Format ▾	Contains Keywords ▾	MARC21 ▾	

Clear form

Will now rely on tag sync national catalog for digital resources, instead of the OA access rule



# Consolidating MARC records

Not quite there!  
Still working on it



University of Notre Dame, Queensland University of  
Technology & Macquarie University for sharing your  
normalisation rules with me.

## Goals

Minimise the number of harvest streams - \$\$\$  
Move to a single NUC symbol instead of one for each  
library - simplify UX & requesting from us

## Implications

Normalisation rules cannot be avoided any more  
because...

- We want a single holding statement even if the title is in multiple locations
- Physical records can be a little messy with local tags - especially old ones that have been migrated to Alma

## Learnings

It takes longer than you think  
There are many possible approaches to preparing your records - recommend breaking the changes into parts eg. Bib data then Holdings data  
Not all libraries are contributing holdings in 984 tags - though you might be told to do that



# Harvesting

---

# Failing with Figshare

[David Witteveen](#)

Coordinator, Library Discovery Platforms  
La Trobe University Library



# The Project

## GOAL

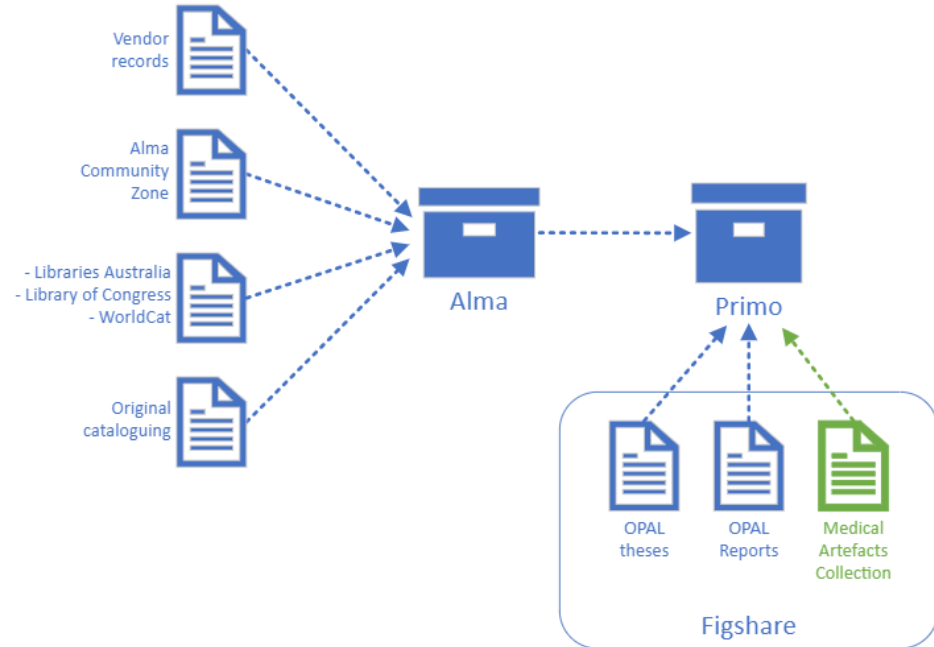
Make the Medical Artefacts Collection discoverable in Primo.

## PROBLEM

This was all brand new to me.

I'd never set up a pipe in Primo Back Office.

And I'd heard about OAI-PMH, but never configured it.



# OPAL - a Figshare repository

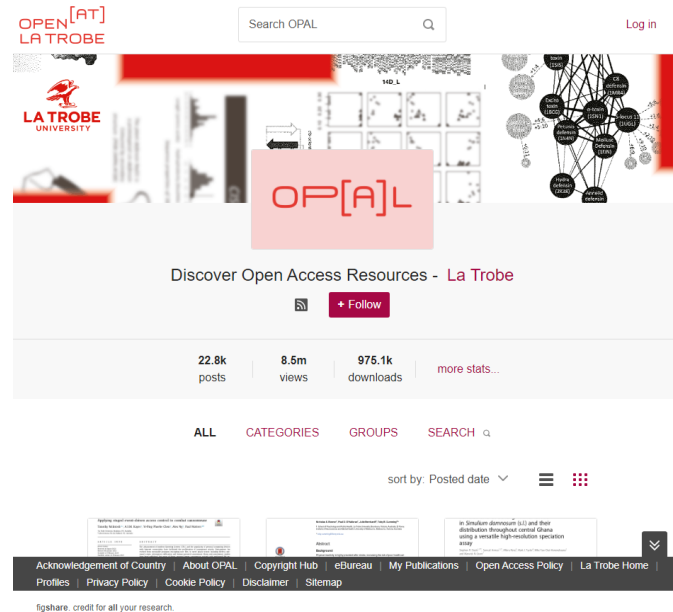
Open Access Resources at La Trobe  
<https://opal.latrobe.edu.au>

Built on [Figshare](#)

Contains:

- Theses
- Reports
- Open publications
- Open data

We import the **theses** and **reports** into Primo to improve discoverability



# The Medical Artefacts Collection

Added to OPAL in June 2022.

“The Medical Artefacts Collection contains a broad range of Artefacts from past medical courses run by La Trobe University and its antecedent institutions. The collection dates from approximately 1915-1990, with the majority of materials likely dating from the 1950s, 1960s, and 1970s.”

Currently 24 items.

Wouldn't it be great if you could find them in Primo?

24 results found

sort by: Date added



**Glass syringe in light blue box, labelled 'Interchangeable ...**  
Physical object posted on 2022-06-24

[University Archives](#)



**"Solila" hypodermic needles in a blue box; and pink box.**  
Physical object posted on 2022-06-24

[University Archives](#)



**Surgical Hypodermic Needles**  
Physical object posted on 2022-06-24

[University Archives](#)



**Needles in M Fabian box**  
Physical object posted on 2022-06-24

[University Archives](#)



**"Beacon" box with hypodermic needle points.**  
Physical object posted on 2022-06-24

[University Archives](#)



**Everett Syringe with needle.**  
Physical object posted on 2022-06-24

[University Archives](#)



**Palmer injector with syringe.**  
Physical object posted on 2022-06-24

[University Archives](#)



**Otoscope, battery operated.**  
Physical object posted on 2022-06-24

[University Archives](#)



# OAI-PMH Sets

Primo lets you harvest from OAI repositories by set.

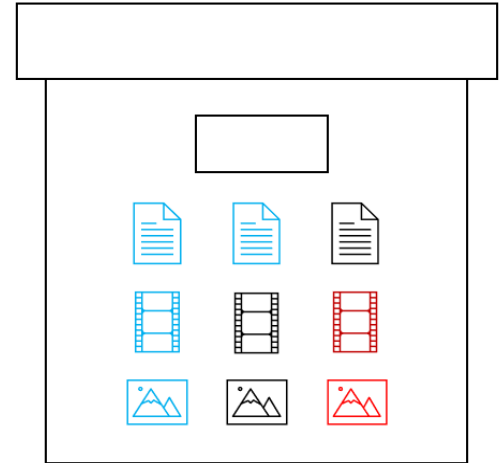
“A **set** is an optional construct for grouping items for the purpose of selective harvesting.”

- OAI-PMH Specification

Which sets are available are defined by the OAI repository.

Example from the diagram:

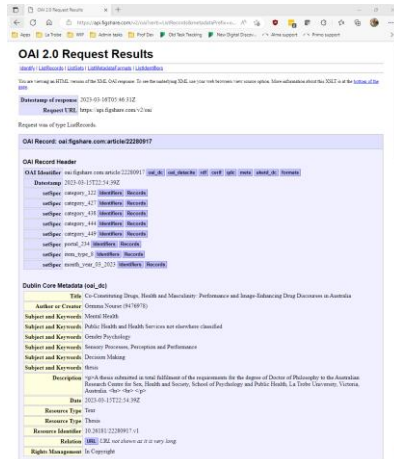
- Every **text** file
- Every **blue** item



# Fun fact!

OAI-PMH sends requests via HTTPS, so you can test things in your browser.

[https://api.figshare.com/v2/oai?verb=ListRecords&metadataPrefix=oai\\_dc&set=portal\\_234&set=item\\_type\\_8](https://api.figshare.com/v2/oai?verb=ListRecords&metadataPrefix=oai_dc&set=portal_234&set=item_type_8)



The screenshot displays a web browser window with the address bar showing the URL: `https://api.figshare.com/v2/oai?verb=ListRecords&metadataPrefix=oai_dc&set=portal_234&set=item_type_8`. The page content is titled "OAI 2.0 Request Results" and includes a sub-header "OAI Record: oai:figshare.com:article2228917".

**OAI Record Header**

OAI Identifier: oai:figshare.com:article2228917 [URL](#) [URL](#) [URL](#) [URL](#) [URL](#) [URL](#) [URL](#) [URL](#) [URL](#) [URL](#)

**Description:** 2023-03-23T12:24:24Z

- oai\_dc: oai\_dc:1 [View Record](#) [Download](#)
- oai\_dc: oai\_dc:47 [View Record](#) [Download](#)
- oai\_dc: oai\_dc:48 [View Record](#) [Download](#)
- oai\_dc: oai\_dc:44 [View Record](#) [Download](#)
- oai\_dc: oai\_dc:49 [View Record](#) [Download](#)
- oai\_dc: oai\_dc:214 [View Record](#) [Download](#)
- oai\_dc: oai\_dc:2 [View Record](#) [Download](#)
- oai\_dc: oai\_dc:201 [View Record](#) [Download](#)

**Dublin Core Metadata (oai\_dc)**

**Title:** Co-Cocaine/ing Drugs, Health and Mortality: Performance and Image-Enhancing Drug Discussion in Australia

**Author or Creator:** Connor, Nicole [614918]

**Subject and Keywords:** Mental Health

**Subject and Keywords:** Public Health and Health Services not elsewhere classified

**Subject and Keywords:** Cocaine Psychology

**Subject and Keywords:** Recovery Processes, Perception and Performance

**Subject and Keywords:** Devises Making

**Subject and Keywords:** Abuse

**Description:** ng-A-Basic submitted as total fulfillment of the requirements for the degree of Doctor of Philosophy in the Australian Research Centre for Sex, Health and Society, School of Psychology and Public Health, La Trobe University, Victoria, Australia. http://dx.doi.org/10.21203/3.11712134.002

**Issue:** 2023-03-23T12:24:24Z

**Resource Type:** Text

**Resource Type:** Thesis

**Resource Identifier:** 10.21203/3.11712134.002

**Relation:** [ORCID](#) [ORCID](#) not shown as it is very long

**English Message:** In Copyright

# How we import **theses** into Primo

In Primo Back Office, we have a pipe called **OPAL\_theses**

Data Source:	OPALtheses
Harvesting method:	OAI
Server:	https://api.figshare.com/v2/oai
Metadata format:	oai_dc
Set:	portal_234&set=item_type_8

**Portal\_234** = La Trobe Uni items  
**set=item\_type\_8** = theses, which makes this easy

The screenshot displays the Primo Back Office interface for configuring a pipe. The page title is "APAC MT - Production" with a server time of "06/03/23 07:31:55 (Server Time)". The breadcrumb trail is "Primo Home > Monitor Primo Status > Pipe Monitoring". The user is identified as "Hello david".

The main content area is titled "Search Engine" and "Define Pipe". The pipe configuration is as follows:

- Owner: La Trobe
- Pipe Name: OPAL\_theses
- Pipe Type: Regular
- Pipe Description: Imports thesis records from OPAL.

The configuration is divided into three sections:

- Source:** Data Source: OPALtheses; Harvesting method: OAI; Server: https://api.figshare.com; Metadata format: oai\_dc; Set: portal\_234&set=item\_type\_8; Encode Resumption Token: false; Start harvesting files/records from: Aug 01, 2020; Start time: 13:21:36.
- Normalization and Enrichment:** Normalization Mapping Set: OPALtheses; Enrichment Set: No Enrichments - Template; Harvested File Format: tar.gz; System Last Stage: FRBR.
- Errors:** Maximum error threshold: 5% (1-100).

Buttons for "Test Connection", "Back to Pipes List", and "Save" are visible at the bottom.



## Medical Artefacts Collection should be similar, right?

All we need to do is define a **set** that selects the items in the Medical Artefact Collection!

Let's look at the [Figshare documentation](#) for OAI-PMH sets:

*At this moment selective harvesting can be performed using sets representing:*

- *an item type*
- *a portal/institution*
- *a category*
- *or a month\_year*

You can't select by **collection**. ☹️

Figshare support suggested we use their API instead. That's where I gave up.

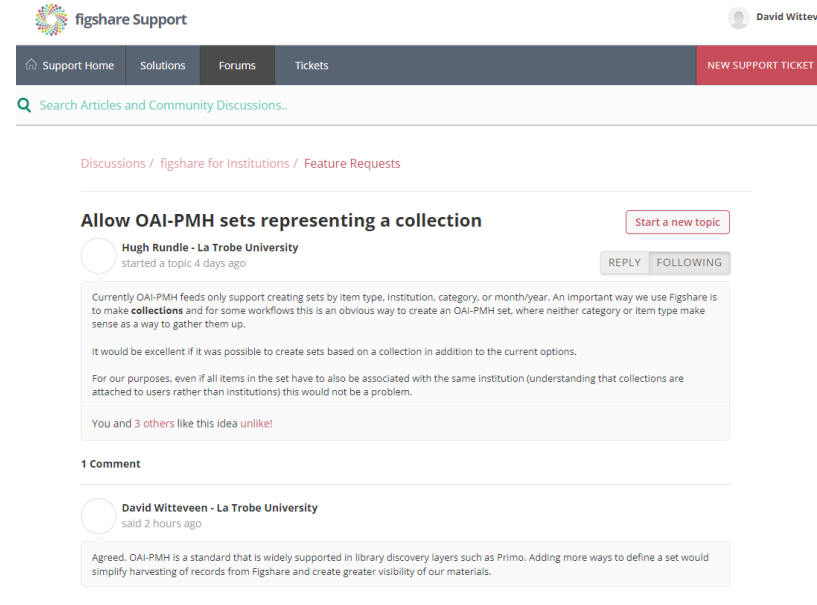


# What now?

We've placed [a request to include sets representing a collection](#) with the Figshare feature requests.

(Support this if you use Figshare!)

And we've asked someone in La Trobe's IT team to investigate the API.



The screenshot shows a forum post on the Figshare Support website. The page header includes the Figshare logo and 'figshare Support', a navigation menu with 'Support Home', 'Solutions', 'Forums', and 'Tickets', and a search bar. The user 'David Witteveen' is visible in the top right. The forum thread is titled 'Allow OAI-PMH sets representing a collection' and was started by 'Hugh Rundle - La Trobe University' 4 days ago. The post content discusses the current limitations of OAI-PMH feeds and proposes creating sets based on collections. It mentions that 3 others like the idea. There is one comment from 'David Witteveen - La Trobe University' 2 hours ago, which agrees with the proposal.

figshare Support

Support Home Solutions Forums Tickets

Search Articles and Community Discussions.

Discussions / figshare for Institutions / Feature Requests

### Allow OAI-PMH sets representing a collection

Hugh Rundle - La Trobe University started a topic 4 days ago

Currently OAI-PMH feeds only support creating sets by item type, institution, category, or month/year. An important way we use Figshare is to make **collections** and for some workflows this is an obvious way to create an OAI-PMH set, where neither category or item type make sense as a way to gather them up.

It would be excellent if it was possible to create sets based on a collection in addition to the current options.

For our purposes, even if all items in the set have to also be associated with the same institution (understanding that collections are attached to users rather than institutions) this would not be a problem.

You and 3 others like this idea [unlike!](#)

1 Comment

David Witteveen - La Trobe University said 2 hours ago

Agreed. OAI-PMH is a standard that is widely supported in library discovery layers such as Primo. Adding more ways to define a set would simplify harvesting of records from Figshare and create greater visibility of our materials.



# Summary

This is a quick summary of about 3 months of research, testing, swearing, and confusion.

I had to learn a lot.

It took me a long time to work out Figshare couldn't do what we wanted, rather than that I just didn't know how to configure things.

## CONCLUSION

This project failed. But it taught me a lot about OAI-PMH for our Trove publishing.

---

# Reaping the Benefits

[Mike Kmiec](#)

Library User Experience Manager  
Western Sydney University



# About the Whitlam Prime Ministerial Collection

Owned by the Whitlam Institute; the University Library is caretaker

Over 35,000 items in the Collection

Digitised and born-digital items are managed and maintained within Rosetta

Rosetta holds over 21,000 intellectual entities (across 2,650 'collections')

POLICY/RESEARCH CULTURE/HERITAGE CIVICS/EDUCATION PM COLLECTION

Whitlam Institute  
WITHIN WESTERN SYDNEY UNIVERSITY

Home About Support What's On Contact Search Store

GOUGH WHITLAM  
PARLIAMENT HOUSE  
BARTON ACT

## Whitlam Prime Ministerial Collection

### Explore the Collection

The Whitlam Prime Ministerial Collection contains more than 35,000 wonderful objects that capture the essence of Gough Whitlam's personality, his government, his public life, and his times. The collection is constantly growing as we attract new treasures that help us continually better understand Australia's greatest reforming Prime Minister.

#### COLLECTION HIGHLIGHTS



#### 1972 'My Time' Speech

This is the original copy of perhaps the most important speech of Gough Whitlam's political career: his 1972 election policy speech. The speech laid out the guiding principles of the Whitlam Government and defined the policy program it intended to enact. The speech contained around 200 specific promises. Whitlam regarded his speech as the contract between his government and the Australian people, and referred to it as his "contract" to the nation. Opening lines: "Men and Women of Australia" will accompany him from the variety of meetings, labor union rallies, club dinners. His copy of the speech



#### Wedding Photograph

Edward Gough Whitlam married Margaret Elizabeth Stowe in 1942, and this is their wedding portrait from that day. The ceremony took place at St Andrew's Church in Brisbane. They were married for almost 70 years until Margaret's death in March 2012.



#### Letter of Dismissal

One of the most significant documents in Australian political history. This is the original letter signed by Gough Whitlam to Edward Stowe of St John's Inn, terminating his commission as Prime Minister, and directing him to government from office. Whitlam had called on the Governor-General on 11 November 1975, the requested to advise him to call a double-dissolution election to break the deadlock that was blocking the passage of his budget. 166 years later, 1969, from down on Gough's desk when Whitlam walked over his office. Gough Whitlam would advise the

SEARCH THE COLLECTION

#### DONATE TO THE COLLECTION

Do you have a story to tell about Gough Whitlam or the Whitlam Government?

Whether you have items to donate or a personal history you'd like to record, we'd love to hear from you.

[Get in touch](#)

#### VISIT A CHANGING AUSTRALIA: THE TIME OF GOUGH WHITLAM

Visit our on-site exhibition to explore the dramatic and transformative years of the Whitlam Government and see the many Whitlam Prime Ministerial Collection items on display.

#### ACCESS THE COLLECTION ONLINE

The Whitlam Prime Ministerial Collection has been largely digitised - which means you can search for everything from family photographs to original speeches with Gough Whitlam's name in the margin.

The collection is housed at the Whitlam Institute and available for research by appointment.

[Get in touch](#)



# Discovery, Previously

**Primo Back Office**

Welcome to Primo Back Office - monitor and configure your settings

**Monitoring & Maintenance**

- > **Monitor Primo Status**  
Stay in control with one glance of all aspects of Primo performance
- > **Primo Utilities**  
Cleanup procedures and More
- > **Primo Analytics**  
View and Design reports
- > **Schedule Primo Analytics**  
Schedule reports and dashboards
- > **Primo Reports**  
View available reports or catalog a new report
- > **Schedule Tasks**  
Schedule Primo tasks such as harvesting, reloading indexes and more
- > **PNX Viewer**  
Browse PNX files

**Configuration & Management wizards**

- > **Ongoing Configuration Wizards**  
Configure and manage the Staff, Institutions, Views, Data sources and Pipes settings to maintain the proper function of Primo
- > **Advanced Configuration**  
Manage the advanced features of Primo, like Search Engine, Installation, UI and PNX

**Total Searches**

40
35
30
25
20
15
10
5
0

**Pipes:**

> De  
De

The Collection was made discoverable through Primo through a separate view

This involved Primo harvesting the Rosetta OAI-PMH feed, which couldn't be migrated

When the time came to move to Primo VE in early 2022...

**footage not found**



# Discovery Import Profiles

Setting this up seemed straightforward

Until it wasn't

< Import Profile Details

Profile Details   Normalization   Delivery

### Profile Details

Profile name	WPMC Rosetta to Primo VE discovery		
Profile description	Harvest WPMC Rosetta files into Primo VE for discovery		
Data Source Code	WPMC_PVE	Data Source Label	Whitlam Prime Ministerial Collection
Originating system	Other	File name patterns	-
Import Protocol	OAI		
Physical source format	XML		
Source format	Dublin Core	Target format	Dublin Core
Status	Active		

### Scheduling

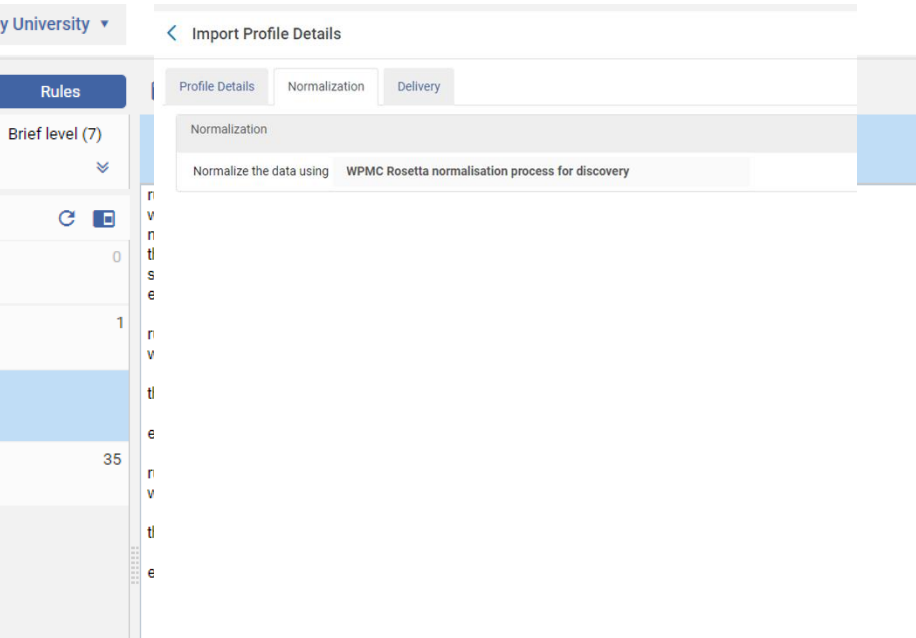
Scheduler status	Active
Scheduling	Every day at 06:00
	<a href="#">Email Notifications</a>

### OAI Details

OAI Base URL	https://rosetta.westernsydney.edu.au/oaiprovider/request		
Authentication	No		
Repository Name	Rosetta	Earliest Date Stamp	13/09/2021
Granularity	YYYY-MM-DDThh:mm:ssZ	Admin EMail/s	
Metadata Prefix	rosetta_dc		
Set	primo (primo)		



# Discovery Import Profiles, Continued



University ▾

< Import Profile Details

Profile Details Normalization Delivery

Normalization

Normalize the data using WPMC Rosetta normalisation process for discovery

Rules

Brief level (7)

0

1

35

Had to create a Normalisation Process Task

Which meant setting up a Normalisation Rule for an External Data Source

Once finished, testing, and lots of it

# Mapping the Metadata

Description is done in Rosetta, using  
dcterms:MediaType for medium

Had to map these to Resource Types in  
Primo VE, for facets

Out-of-the-box Resource Types weren't  
exactly correct

	A	B	C	D
artwork		ephemera (membership card)	object (mixed material)	photograph (colour)
artwork (bark painting)		ephemera (menu)	object (paper)	photograph (digital media)
artwork (t...				isc)
artwork (c...				microfiche)
artwork (e...	1	Archival Material/Manuscript	Map	tereograph)
artwork (l...	2	Article	Market Research	
artwork (r...	3	Audio	Microform	(digital med
artwork (r...	4	Book	Newsletter article	(file)
artwork (r...	5	Book Chapter	Newspaper	
artwork (r...	6	Book Review	Newspaper article	
artwork (v...	7	Computer File	Other	
audio (dig...	8	Conference proceeding	Patent	
audio (dis...	9	Database	Reference entry	
audio (tap...	10	Dataset	Report	
data (digi...	11	Dissertation	Score	
data (disc...	12	Government document	Standard	
diary	13	Image	Text Resource	
directory	14	Journal	Video	
ephemera	15	Legal Document	Web Resource	
ephemera	16			
ephemera (greeting card)		object (glass)	photograph (b&w)	
ephemera (invitation)		object (leather)	photograph (colour negative)	
ephemera (itinerary)		object (metal)	photograph (colour transparency)	

# Mapping the Metadata, Continued

Could we rationalise and map Rosetta Media Types to extant Primo VE Resource Type?

Could we use Primo VE's Local Resource Types instead?

**Rosetta type**  
**audio (three types)**  
**data (two types)**

Local Resource Types

1 - 3 of 3

	Enabled	Code
1	<input checked="" type="checkbox"/>	exam_papers
2	<input checked="" type="checkbox"/>	learning_guides
3	<input checked="" type="checkbox"/>	open_educational_res

## Content Type ^

- paper\_file (13)
- paper\_map (12)
- monograph (10)
- paper\_document (8)
- ephemera (5)
- ephemera\_invitation (2)
- ephemera\_itinerary (2)
- press\_clipping (2)
- audio\_disc (1)
- diary (1)
- ephemera\_program (1)
- object\_fabric (1)
- object\_metal (1)
- object\_paper (1)
- object\_silver (1)
- photograph\_album (1)
- photograph\_microfiche (1)

**Primo VE type**

Dataset

- exam\_papers
- learning\_guides
- open\_educational\_resources



## Outcomes



Whitlam Prime Ministerial Collection now in Primo VE

Also findable in the Library's instance of Primo VE

Learnt much about harvesting and its limitations

Opened opportunities for how we work with metadata

Reminder to document configuration decisions



# Publishing





# OAI-PMH publishing and harvesting at UNSW

Bronwyn King, Digital Library Initiatives, Library

1. Alma publishing to RapidILL
2. Recollect publishing to Primo (back office)

We also:

1. Publish Alma journal holdings to Third Iron's LibKey and BrowZine services
2. Have started process to replace RIS exports to Libraries Australia with OAI-PMH to Trove



# Publishing from Alma to RapidILL

RapidILL is an interlibrary loans system integrated with Alma's Resource Sharing management system. In order for Rapid to know what UNSW holds, publishing of our holdings to their database is necessary.

## Goals:

- Replace data exports with OAI-PMH publishing for print journals and eJournals. Achieved: 2018
- Setup OAI-PMH publishing for eJournals taking licensing into consideration. Achieved: 2018
- Setup OAI-PMH publishing for eBooks taking licensing into consideration. Achieved: 2022

Six publishing sets were identified:

1. Print journals
2. Print books
3. eJournals - "LocalOnly"
4. eJournals - "LendableInternational"
5. eBooks - "LocalOnly"
6. eJournals - "LendableInternational"



# Publishing to RapidILL - Print process

Process for each print set:

1. Create Alma logical set
2. Create Alma publishing profile
3. Publish sample of records
4. Open Salesforce case and liaised with Ex Libris to complete setup on both sides
5. Conduct initial full publishing
6. Test and when both sides agree, schedule weekly incremental runs



# Print books - 1. Created Alma logical set

Criteria:

Search in: **Physical titles** ▾ Zone:  **Institution**  Community

Title: Bibliographic Level	▾	Equals	▾	Monograph/item	▾
<b>AND</b> OR	ISBN	▾	Contains Keywords	▾	*
<b>AND</b> OR	Tag Suppressed (Title)	▾	Equals	▾	No
<b>AND</b> OR	Library (Holdings)	▾	Equals	▾	Academy Library, UNSW Canberra <b>×</b> Law Library <b>×</b> Library Storage <b>×</b> Main Library <b>×</b> Paddington Library <b>×</b>

# Print books - 2. Created Alma publishing profile

Profile Details | Data Enrichment

Profile Details

Profile Name \* **RapidILL publishing - print books**

Profile Description

Publishing Parameters

Status  Active  Inactive

Scheduling **Every Sunday at 09:00** [Edit Scheduling](#)

[Email Notifications](#)

Content

Set name \* **LASU - RapidILL - print books**

Additional set name [Select from a list](#)

Filter out the data using

Publish on: **Bibliographic level**

Output format **MARC21 Bibliographic**

Publishing Protocol

Profile Details | **Data Enrichment**

Physical Holdings Enrichment

Add Holdings information

Holdings Tag	Holdings Subfield	Bib Tag	Bib Subfield
852	b	852	b
852	c	852	c
852	h	852	h
852	i	852	i
852	k	852	k
852	m	852	m

ADDITIONAL HOLDINGS FIELDS

Created by field  Subfield

Create date field  Subfield

Updated by field  Subfield

Update date field  Subfield

Library name field  Subfield

Location name field  Subfield

Suppress from publishing field  Subfield

Exclude suppressed records



# Publishing to RapidILL - Electronic resources

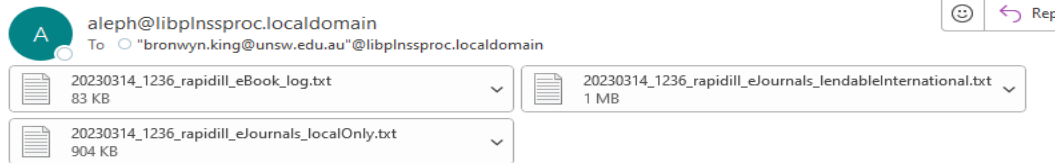
- Semi-manual process run 3 times per year before the beginning of each term
- Process must determine which resources are available to be “lent” to all RapidILL partners including overseas and which cannot be lent and create new Alma sets for them
- Setup process:
  - a. Create MMS ID lists for the “LendableInternational” and “LocalOnly” cases using Alma Analytics.
  - b. Create itemized sets using the lists created under step 1.
  - c. Create Alma publishing profiles
  - d. Conduct initial full publishing
  - e. Open Salesforce case and liaised with Ex Libris to complete setup on UNSW’ and their sides
  - f. For the ongoing semi-manual incremental runs, create a script to automatically run the Alma analytics reports, prepare input files for the creation of Alma sets and email the relevant member of staff



# Publishing to RapidILL - Incremental runs script

## Output from eJournals script run:

### RapidILL eJournals MMS IDs lists



aleph@libplnssproc.localdomain  
To: "bronwyn.king@unsw.edu.au"@libplnssproc.localdomain

20230314\_1236\_rapidill\_eBook\_log.txt  
83 KB

20230314\_1236\_rapidill\_eJournals\_localOnly.txt  
904 KB

20230314\_1236\_rapidill\_eJournals\_lendableInternational.txt  
1 MB

The run\_rapidill\_eresources\_analytics.sh job has completed. The MMS ID count is:

- eJournals Lendable count: 83485
- eJournals LocalOnly count: 51381

See the attached the log file and output (three MMS ID lists for the creation of Alma itemised sets).

If the MMS ID lists were created the next steps are:

1. Backup the current Alma sets:
  - DLI - RapidILL - eJournals - LendableInternational
  - DLI - RapidILL - eJournals - LocalOnly

By renaming each Alma set to include the original creation date.  
e.g.



# Publishing to RapidILL

## Conclusions:

1. Print resources setup is very straight forward. No major issues since initial setup
1. Electronic resources more complex due to the need to create “LendableInternational” and “LocalOnly” publishing sets for each publishing run. In theory entire process could be automated using Alma set APIs but may be more trouble than it is worth.
1. Looking to adding “Australia only” sets for eBooks and eJournals to reduce the incoming requests rejections





## Recollect OAI-PMH publishing to Primo

UNSW uses [Recollect](#) to store, make accessible and preserve our University Archives Photographs, student publications such as Tharunka and other special digital collections. To enhance the collections' discoverability it was decided that the open access content be published to Primo.

Process:

1. Analyse Recollect metadata and develop publishing specs
2. Opened a ticket with vendor requesting publishing setup, based on specs. Also, the white listing of our Primo environments was requested.
3. Vendor configured OAI-PMH publishing based on UNSW specs
4. In the Primo back office, UNSW setup configuration to harvest the publishing
5. Scheduled daily harvests in Primo back office



# Recollect OAI-PMH publishing to Primo

Field mapping specs:

Recollect item template field	OAI-PMH field
About	dc:description
Author	dc:creator
Choreographer	dc:creator
Cinematographer	dc:creator
Contact us	dc:description
Contributor:	dc:contributor
Copyright	dc:rights
Creator	dc:creator
Dancers	dc:creator
Date	dc:date
Date taken	dc:date
Description	dc:description
Director	dc:creator
Editor	dc:contributor
Illustrator	dc:contributor
Note	dc:description
Organisation	dc:subject
Photographer	dc:creator
Production company	dc:subject
Publication date	dc:date
Publisher	dc:publisher
Subject	dc:subject
Title	dc:title
	dc:type



# Recollect OAI-PMH publishing to Primo

Primo harvesting:

Owner: UNSW Alma    Pipe Name: UNSW\_RECOLLECT    Pipe Type: Regular

Pipe Description:

## Source

Data Source: \*

UNSW\_RECOLLECT ▼

Harvesting method: \*

OAI ▼

Server: \*

<https://digitalcollectio>

Metadata format: \*

oai\_dc

Set:

Encode Resumption Token:

false ▼

Start harvesting files/records from: \*

## Normalization and Enrichment

Normalization Mapping Set: \*

UNSW\_RECOLLECT ▼

Enrichment Set: \*

No Enrichments ▼

Harvested File Format: \*

\*.tar.gz ▼

System Last Stage: \*

FRBR ▼

Primo display of harvested content:

3



STREAMING VIDEO

[Dance Royale](#)

George Vaviloussakis

2007

The ReelDance Archive

[Check online access](#) [↗](#) [▶](#)

4



STREAMING VIDEO

[Frankfurt dance cuts](#)

Prue Lang; Alan Barnes; Jone San Martin; Vanessa LeMat; Lutz Gregor

2004

The ReelDance Archive

[Check online access](#) [↗](#) [▶](#)

5



STREAMING VIDEO

[How to dance if you can't](#)

Tamara Kapp ; Tamara Kapp

2005

The ReelDance Archive



## Recollect OAI-PMH publishing to Primo

### Conclusions:

1. Publishing specifications development fairly straight forward
2. Vendor very responsive and was able to set publishing quickly
3. Back office Primo setup quite complex

### Next steps:

1. Publish content to Trove
2. Publish content to Primo VE (VE go live = Sep 2023)



# Publishing OAI from Alma for discovery and preservation

Rich Robertson

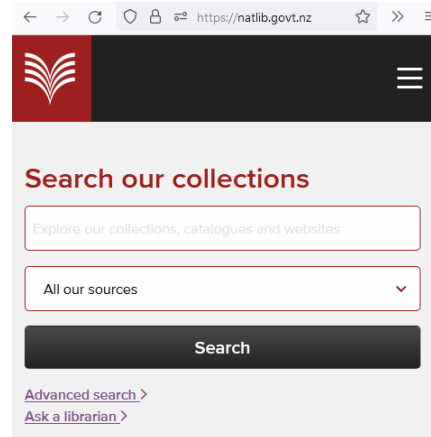
Metadata workflow specialist, Collection description

Te Puna Mātauranga o Aotearoa - National Library of New Zealand

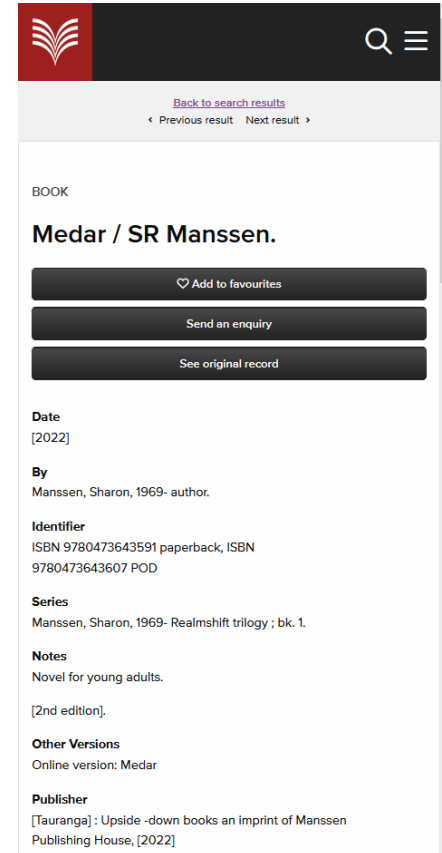
[richard.robertson@dia.govt.nz](mailto:richard.robertson@dia.govt.nz)

# Alma publishing to NatLib search

- NatLib search is the main discovery search found on the National Library website: <https://natlib.govt.nz/>
- Aggregates metadata from published collections, unpublished collections, PapersPast and more.
- Provides a starting point for users searching our collections.
- Natlib records are also indexed by search engines.
- Managed by Digital Experience team.



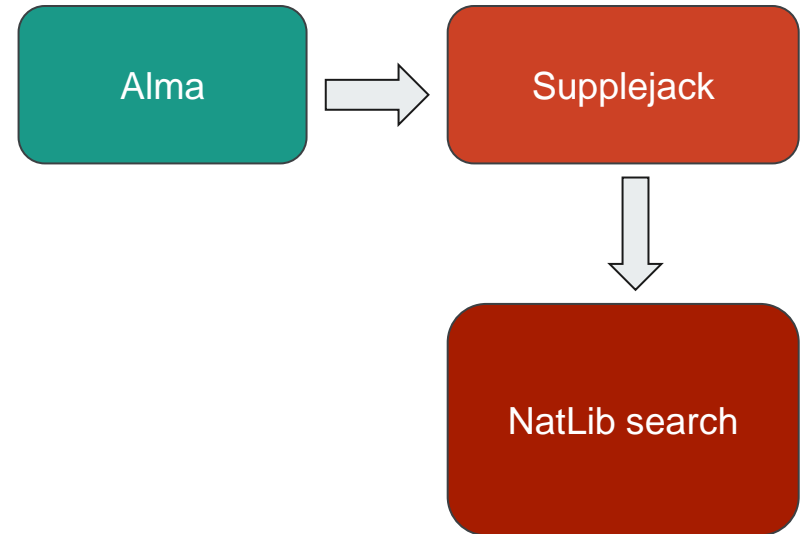
The screenshot shows the NatLib search homepage. At the top is a navigation bar with the NatLib logo on the left and a search icon on the right. Below the navigation bar is a search box with the text "Search our collections" in red. The search box contains the placeholder text "Explore our collections, catalogues and websites". Below the search box is a dropdown menu with the text "All our sources" and a downward arrow. Below the dropdown menu is a black "Search" button. At the bottom of the search area are two links: "Advanced search >" and "Ask a librarian >".



The screenshot shows the search results page for "Medar / SR Manssen.". At the top is a navigation bar with the NatLib logo on the left and a search icon on the right. Below the navigation bar is a search bar with the text "Back to search results" and "Previous result Next result >". Below the search bar is a section titled "BOOK" and "Medar / SR Manssen.". Below this section are three buttons: "Add to favourites", "Send an enquiry", and "See original record". Below the buttons is a section titled "Date" with the text "[2022]". Below this section is a section titled "By" with the text "Manssen, Sharon, 1969- author.". Below this section is a section titled "Identifier" with the text "ISBN 9780473643591 paperback, ISBN 9780473643607 POD". Below this section is a section titled "Series" with the text "Manssen, Sharon, 1969- Realmshift trilogy ; bk. 1.". Below this section is a section titled "Notes" with the text "Novel for young adults. [2nd edition]". Below this section is a section titled "Other Versions" with the text "Online version: Medar". Below this section is a section titled "Publisher" with the text "[Tauranga] : Upside -down books an imprint of Manssen Publishing House, [2022]".

## Alma publishing to NatLib search

- Alma publishing profile publishes MARC records via OAI from set of all unsuppressed records.
- [Supplejack](#) is used to harvest Alma records daily - picking up new, updated and deleted records.
- MARC records are mapped into DigitalNZ metadata standard which is based on Dublin Core with local DigitalNZ attributes.
- Records are exposed using Supplejack API, and indexed into the NatLib search.







## Publish Dublin Core instead of MARC?


Alma does allow [crosswalking of MARC to Dublin Core records](#)

It wasn't feasible at the time, as a lot of important metadata isn't crosswalked, including:

- Publisher information from the 264 field
- Holding information
- Descriptive information from the 300 field (extent, units, dimensions)
- Music identifiers from the 028 field
- Additional subfields in some fields - e.g. 245c for statement of responsibility

**February release:** the crosswalk is now configurable...

```
<record>
  <dc:title>Publishing OAI from Alma
for discovery and preservation
</dc:title>
  <dc:creator>Robertson, Rich
</dc:creator>
  <dc:description>Part of ANZREG
webinar on OAI</dc:description>
  <dc:publisher>ANZREG</dc:publisher>
  <dc:date>March 2023</dc:date>
  <dc:type>Text</dc:type>
  <dc:language>eng</dc:language>
</record>
```



# Additional publishing profile for targeted publishing

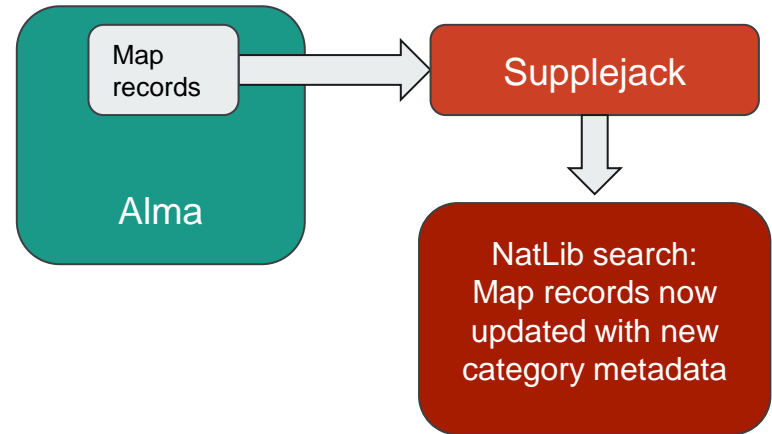
An additional publishing profile is now used for manually publishing a targeted set of records to NatLib search.

NatLib search occasionally needs to update a specific set of Alma records:

- To enable new functionality - for example, a new filter for maps
- To address issues with legacy data

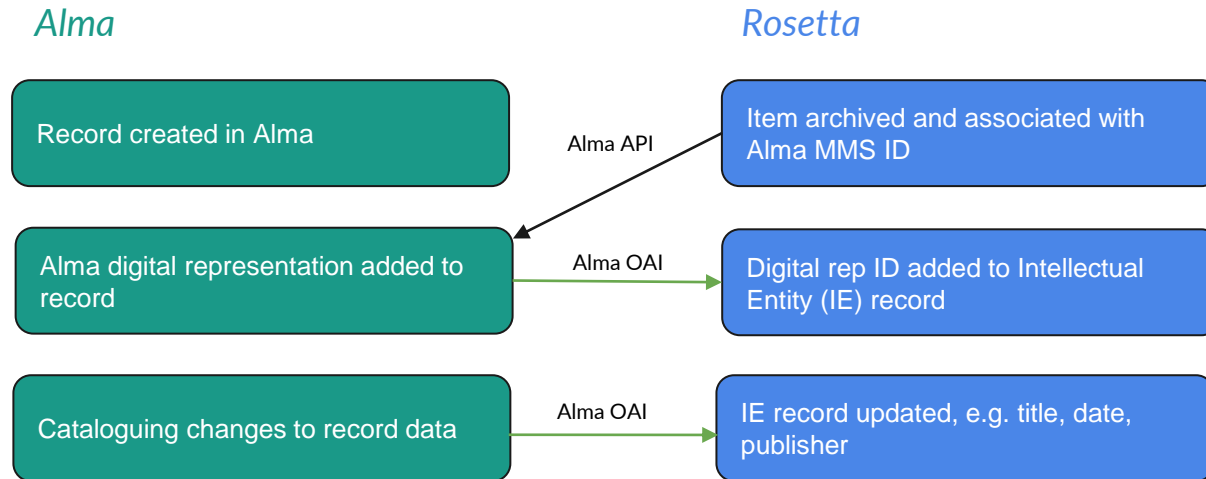
This has been beneficial for both teams - as the process does not impact the regular publishing and harvesting processes.

For example:



# Alma publishing to Rosetta

Alma OAI publishing is part of a bigger digital workflow happening between Alma and Rosetta:





## Other useful resources

Nancy Babb (2022) [Reaping Resources for Recommenders](#) (ENUG) - harvesting Libguides into MarcEdit and transforming for Resource Recommender (Primo)